



Lässt sich erweitertes Zusammenwirken zwischen Menschen und intelligenten Maschinen empirisch feststellen? Überlegungen zur Adaptierung des Turing-Tests*

REINHOLD HAUX

Peter L. Reichertz Institut für Medizinische Informatik der Technischen Universität Braunschweig und der Medizinischen Hochschule Hannover (PLRI), Mühlenpfordtstraße 23, 38106 Braunschweig, reinhold.haux@plri.de, www.plri.de

Zusammenfassung

Die Kommission Synergie und Intelligenz der Braunschweigischen Wissenschaftlichen Gesellschaft befasst sich in grundsätzlicher Weise mit den technischen, ethischen und rechtlichen Herausforderungen des Zusammenwirkens lebender und nicht lebender Entitäten, die im Zeitalter der Digitalisierung möglich geworden sind. Eine Frage, die während der Kommissionsarbeit gestellt wurde, bezog sich auf die Bedeutung des Begriffs des erweiterten Zusammenwirkens, insbesondere der Bedeutung von 'erweitert'. Damit verbunden war die Frage: Kann erweitertes Zusammenwirken zwischen Menschen und intelligenten Maschinen festgestellt werden?

Mit dem SYnENZ-Test wird in dieser Ausarbeitung ein Ansatz vorgeschlagen, mit dem sich empirisch ermitteln lässt, ob ein solches erweitertes Zusammenwirken vorliegt. Der Ansatz greift die Idee des Turing-Tests auf und nutzt bekannte Ansätze der Versuchsplanung und des statistischen Testens. Der Test wird am Beispiel des Anwendungsfalls Behandlung von Patienten erläutert. In dem SYnENZ-Test gibt es drei Entitätsgruppen: (1) Patienten, die Behandlung benötigen. (2) Behandelnde, die Behandlung erbringen; diese sind eingeteilt in zwei Behandlungsteams, die zum einen aus Fachkräften in Gesundheitsberufen und zum anderen aus solchen Fachkräften gemeinsam mit intelligenten Maschinen bestehen. (3) Fragesteller, die den Behandlungserfolg ermitteln und analysieren. Sollten die Fragesteller nach Durchführung einer dafür adäquat geplanten Studie anhand des SYnENZ-Tests zu der Aussage kommen, dass der Behandlungserfolg in dem Behandlungsteam mit Fachkräften in Gesundheitsberufen und intelligenten Maschinen höher ist, als in dem Behandlungsteam mit Fachkräften in Gesundheitsberufen, dann sollte von einem erweiterten Zusammenwirken ausgegangen werden.

Der vorgeschlagene Ansatz wird diskutiert, auch im Hinblick auf Erweiterungen über den hier beschriebenen Anwendungsfall hinaus. Nicht zuletzt wird empfohlen entsprechende Studien mit SYnENZ-Tests durchzuführen, da erfahrungsgemäß solche Studien der Versachlichung und Konkretisierung bei der Behandlung von Problemen dienen, auch bei dem Thema der technischen, ethischen und die rechtlichen Herausforderungen des

* Schriftliche Ausarbeitung von zwei Vorträgen, die zu dieser Thematik am 16.7.2020 und am 4.2.2021 auf Zirkeln der Kommission Synergie und Intelligenz (SYnENZ) der Braunschweigischen Wissenschaftlichen Gesellschaft (BWG) gehalten wurden. Der Autor war von 2017 bis 2021 Sprecher der SYnENZ-Kommission der BWG. Eine englische Ausarbeitung ist geplant, Arbeitstitel: 'SYnENCE Test: Adapting the Turing Test idea to measure extended collaboration of humans and intelligent machines'.



Zusammenwirkens lebender und nicht lebender Entitäten, die im Zeitalter der Digitalisierung möglich geworden sind.

Stichwörter: Turing-Test, SYnENZ-Test, Zusammenwirken; natürliche Intelligenz; menschliche Intelligenz; künstliche Intelligenz; Evaluation; randomisierte Studien.

1 Kann erweitertes Zusammenwirken zwischen Menschen und intelligenten Maschinen festgestellt werden?

1.1 Zielsetzung und Leitfragen der BWG-Kommission Synergie und Intelligenz

Die Kommission Synergie und Intelligenz (SYnENZ, [1]) der Braunschweigischen Wissenschaftlichen Gesellschaft (BWG, [2]) befasst sich in grundsätzlicher Weise mit den technischen, ethischen und rechtlichen Herausforderungen des Zusammenwirkens lebender und nicht lebender *Entitäten*, die im Zeitalter der Digitalisierung möglich geworden sind. Dabei soll es nicht nur um neue Kommunikations- und Interaktionsformen zwischen Menschen gehen. Auch das erweiterte Zusammenwirken mit 'nicht lebenden' Objekten (Maschinen, ...) oder mit anderen Lebewesen (Tieren, Pflanzen, ...), in welcher Art auch immer, ist Gegenstand der Kommissionsarbeit ([3]). Den Rahmen der Kommissionsarbeit bilden die folgenden fünf Leitfragen ([1], [3]):

- 1 Wie wird unser Zusammenleben in Zeiten zunehmender Digitalisierung in Zukunft aussehen?
- 2 Welche Synergien ergeben sich durch das nun mögliche erweiterte Zusammenwirken von Menschen, Tieren und Pflanzen einerseits und von Maschinen andererseits, d.h. von natürlicher und künstlicher Intelligenz?
- 3 Können wir zwischen bloß zeitgemäßen und angemessenen Formen des Zusammenwirkens unterscheiden?
- 4 Und können wir Empfehlungen geben, um angemessene Formen zu erreichen und um zwar zeitgemäße, aber problematische Formen zu vermeiden?
- 5 Lässt sich ein Grad des Zusammenwirkens bestimmen?

1.2 Die Frage des erweiterten Zusammenwirkens

Eine Frage, die während der Kommissionsarbeit mehrfach an den Verfasser dieser Ausarbeitung gestellt wurde, bezog sich auf die Bedeutung des Begriffs des erweiterten Zusammenwirkens, insbesondere der Bedeutung von 'erweitert'. Damit verbunden war die Frage, ob erweitertes Zusammenwirken denn ermittelt werden könne.

Häufig bezog sich diese Frage auf das erweiterte Zusammenwirken zweier bestimmter Entitäten, dem erweiterten Zusammenwirken zwischen Menschen einerseits und *intelligenten Maschinen* andererseits. Als intelligente Maschinen werden hier solche Maschinen bezeichnet, deren Funktionalität so umfangreich ist, dass diese Maschinen komplexe Entscheidungen, wie beispielsweise in der Diagnostik oder wie bei Therapien, treffen können oder zumindest maßgeblich an solchen Entscheidungen mit beteiligt sind. Solche Maschinen sind üblicherweise realisiert als auf Computern implementierte Programme.

Für weitere Erläuterungen dazu sei auf [4] verwiesen. Die maßgebliche Mitbeteiligung von Maschinen an solchen Entscheidungen wird in der Literatur auch mit dem Begriff geteilte Autonomie (engl.: shared autonomy) oder Team-Maschinen-Interaktion (engl.: team-machine interaction) verbunden ([5], [6]).

Die 'Frage des erweiterten Zusammenwirkens' soll hier folgendermaßen formuliert werden:

F_{SYnENZ}: Kann erweitertes Zusammenwirken zwischen Menschen und intelligenten Maschinen festgestellt werden?

1.3 Ziel dieser Ausarbeitung

In dieser Ausarbeitung wird ein Ansatz vorgeschlagen, mit dem sich – zumindest aus Sicht des Verfassers – empirisch ermitteln lässt, ob ein solches erweitertes Zusammenwirken vorliegt. Der Ansatz greift die Idee des Turing-Tests auf ([7], Erläuterungen folgen dazu in dieser Ausarbeitung in Abschnitt 2) und nutzt bekannte, ‘klassische’ Ansätze der Versuchsplanung und des statistischen Testens (Erläuterungen in dieser Ausarbeitung in Abschnitt 3.4.1).

1.4 Hinweise zu dieser Ausarbeitung

Der einfacheren Lesbarkeit halber sind einige der in dieser Ausarbeitung verwendeten Begriffe am Ende des Textes in einem Abschnitt Begriffe nochmals erläutert, zumal manche Begriffe hier durchaus spezifisch durch den Verfasser definiert sind. Sie hätten auch anders definiert oder verstanden werden können. Im Text wurden diese Begriffe bei der ersten Verwendung kursiv gesetzt.

Dem Verfasser ist es nicht gelungen, personenbezogene Begriffe angemessen geschlechtsneutral zu verwenden ohne dass dadurch die Lesbarkeit des Textes zu sehr durch komplizierte Formulierungen gelitten hätte. Bezeichnungen bei Personen bitte ich geschlechtsneutral sowohl auf Frauen und auf Männer zu beziehen, auch wenn dies nicht immer explizit so formuliert ist.

2 Alan Turgings Frage zu maschineller Intelligenz transponiert auf den Turing-Test

Die Frage die Alan Turing in seiner 1948 erschienenen Arbeit über Rechenmaschinen und Intelligenz eigentlich behandeln haben möchte, lautet

F_{Turing}: Können Maschinen denken?

(engl.: „Can machines think?“ [7], S. 433). Er stellt direkt anschließend fest, dass diese wie auch die für ihn damit zusammenhängende Frage, ob Maschinen intelligent sein können, wohl schwierig zu behandeln sei. Er schlägt deshalb vor, diese Frage zu ersetzen durch eine andere Frage, die mittels eines Imitationsspiels geklärt werden kann (engl.: imitation game, heute häufig als Turing-Test bezeichnet, [7], S. 433-434).

Kurz und vereinfacht zusammengefasst sieht der Turing-Test folgendermaßen aus: Ein (menschlicher) Fragesteller (engl.: interrogator) kommuniziert mit zwei Entitäten auf Distanz mittels Schriftsprache. Eine dieser Entitäten ist ein Mensch, die andere eine Maschine. Die Fragen stellende Person versucht durch ihre Fragen herauszubekommen, wer Mensch und wer Maschine ist. Sie ist insofern unvoreingenommen, als dass sie durch diese Art der Kommunikation die zwei Entitäten weder hören noch sehen und damit nicht unmittelbar erkennen kann, wer Mensch und wer Maschine ist. Kann ein Fragesteller anhand dieser Befragung nicht zweifelsfrei feststellen wer Mensch und wer Maschine ist, dann sollte man diese Maschine als intelligent bezeichnen.



Der von Englisch nach Deutsch übersetzte Originaltext dazu:

Die neue Form des Problems kann in Form eines Spiels beschrieben werden, das wir als 'Imitationsspiel' bezeichnen. Es wird mit drei Personen gespielt, einem Mann (A), einer Frau (B) und einem Fragesteller (C), der von beiden Geschlechtern sein kann. Der Fragesteller hält sich in einem Raum abseits von den anderen beiden auf. Ziel des Spiels ist es, dass der Fragesteller herausfindet, welcher der beiden anderen der Mann und welcher die Frau ist. Er kennt sie durch die Bezeichnungen X und Y, und am Ende des Spiels sagt er entweder 'X ist A und Y ist B' oder 'X ist B und Y ist A'. Der Fragesteller darf also Fragen an A und B stellen:

C: Würde X mir bitte die Länge seiner oder ihrer Haare sagen?

Nehmen wir nun an, X ist tatsächlich A, dann muss A antworten. Das Ziel von A im Spiel ist es, zu versuchen, C zu einer falschen Identifizierung zu veranlassen. Seine Antwort könnte also lauten:

'Mein Haar ist kraus, und die längsten Strähnen sind etwa 15 cm lang.'

Damit der Tonfall dem Fragesteller nicht hilft, sollten die Antworten schriftlich oder besser noch maschinengeschrieben sein. Ideal ist es, wenn ein Fernschreiber zwischen den beiden Räumen kommuniziert. Alternativ können die Fragen und Antworten auch von einem Vermittler wiederholt werden. Das Ziel des Spiels für den dritten Spieler (B) ist es, dem Fragesteller zu helfen. Die beste Strategie für sie ist wahrscheinlich, wahrheitsgemäße Antworten zu geben. Sie kann ihren Antworten Dinge hinzufügen wie 'Ich bin die Frau, hören Sie nicht auf ihn!'; aber es wird nichts nützen, da der Mann ähnliche Bemerkungen machen kann.

Wir stellen nun die Frage: 'Was wird passieren, wenn eine Maschine die Rolle von A in diesem Spiel übernimmt?' Wird der Fragesteller bei einem solchen Spiel genauso oft falsch entscheiden wie bei einem Spiel zwischen einem Mann und einer Frau? Diese Fragen ersetzen unsere ursprüngliche: 'Können Maschinen denken?'...

([7], S. 433-434, übersetzt mittels DeepL, [8]).

Alan Turing hat damit die eingangs gestellte Frage transponiert auf eine andere Frage, die über den Turing-Test beantwortet werden soll:

F_{Turing} (transponiert): Können Maschinen den Turing-Test bestehen? Falls ja, dann kann wohl auch F_{Turing} Können Maschinen denken? mit ja beantwortet werden.

Diese Umsetzung der Frage scheint insofern plausibel zu sein, als dass der Turing-Test leicht zu verstehen ist und dass er bis heute als 'der' – sicherlich auch kritisch diskutierte – Test genannt wird, wenn es um die Prüfung der Frage geht, ob Maschinen als intelligent bezeichnet werden können ([9], [10], [11], Kapitel 3).

3 Der SYNENZ-Test

3.1 Beschränkung auf bestimmte Szenarien erweiterten Zusammenwirkens

Die Frage F_{SYENZ} , also die Frage des erweiterten Zusammenwirkens von Menschen und intelligenten Maschinen, soll nun in ähnlicher Weise transponiert werden auf einen Test, der hier als SYNENZ-Test bezeichnet werden soll.

Neben der Beschränkung auf diese beiden Entitäten – Menschen und intelligente Maschinen, also nicht, wie in Abschnitt 1.1 genannt, lebende und nicht lebende Entitäten – soll noch eine weitere Einschränkung erfolgen. Wir beschränken uns auf das Anwendungsgebiet Medizin und Gesundheitsversorgung und dort auf die *Behandlung von Patienten* als Anwendungsfall. Inwieweit sich der SYNENZ-Test auch auf weitere Entitäten oder Anwendungsgebiete anwenden lässt, wird in Abschnitt 4 besprochen. Da im Folgenden bei Maschinen immer intelligente Maschinen gemeint sind, soll der besseren Lesbarkeit halber auch verkürzt der Begriff *Maschinen* verwendet werden.

Für die Einschränkung auf das Anwendungsgebiet Medizin und Gesundheitsversorgung und dort noch weitergehend auf die Behandlung von Patienten als Anwendungsfall gab es zwei Gründe:



- Zum einen ist der Verfasser Medizininformatiker. Medizin und Gesundheitsversorgung ist 'sein' Anwendungsgebiet, auf dem er seit mehreren Jahrzehnten arbeitet.
- Zum anderen und hier wichtiger ist, dass es sich besonders im Hinblick auf die Behandlung von Patienten um ein Anwendungsgebiet mit spezifischen Eigenschaften handelt und dass diese Eigenschaften für die Konstruktion des SYnENZ-Tests wichtig waren.

Die Behandlung von Patienten soll hier sehr breit definiert werden. Es geht um Maßnahmen der Gesundheitsversorgung, die durch *Behandelnde*, das sind in der Regel Fachkräfte in Gesundheitsberufen (Ärztinnen/Ärzte, Pflegekräfte, ...), für Patienten erbracht werden. Diese Maßnahmen der Gesundheitsversorgung können Maßnahmen der Prävention, der Diagnostik, der Therapie und der Pflege umfassen (eine differenziertere Betrachtung dazu befindet sich in [4], S. 184-187, Abschnitt 18.3 Lebenssituationen). Von Bedeutung ist, dass die Beziehung zwischen Patienten und Fachkräften in Gesundheitsberufen „durchgehend asymmetrisch“ ist ([12], S. 15): Patienten benötigen Maßnahmen der Gesundheitsversorgung und Fachkräfte in Gesundheitsberufen wollen diese Maßnahmen der Gesundheitsversorgung erbringen ([12], S. 15-19).

Die oben bei Behandelnden gemachte Einschränkung „in der Regel“ bezieht darauf, dass die Behandlung von Patienten auch von anderen Entitäten erbracht werden kann. Dies können beispielsweise pflegende Angehörige sein, die häufig einen wichtigen Beitrag in der Gesundheitsversorgung von Patienten leisten ohne selbst Fachkräfte der Gesundheitsversorgung sein zu müssen. In dieser Ausarbeitung soll es jedoch um den Beitrag von Maschinen gehen, dies im Rahmen des erweiterten Zusammenwirkens von Menschen und solchen intelligenten Maschinen.

3.2 Aufbau des SYnENZ-Tests

In dem SYnENZ-Test gibt es drei *Entitätsgruppen*.

Die erste Entitätsgruppe umfasst

(PAT) Patienten $\mathbf{PAT} := \{\mathbf{PAT}_1, \dots, \mathbf{PAT}_{n_{\mathbf{PAT}}}\}$, $n_{\mathbf{PAT}} \geq 2$,

d.h. Menschen, die Behandlung benötigen.

Die zweite Entitätsgruppe umfasst Behandelnde. Diese sind eingeteilt in zwei *Behandlungsteams*,

- einem Behandlungsteam (HCP)

mit Fachkräften in Gesundheitsberufen (engl.: health care professionals)

$\mathbf{HCP} := \{\mathbf{HCP}_1, \dots, \mathbf{HCP}_{n_{\mathbf{HCP}}}\}$, $n_{\mathbf{HCP}} \geq 1$,

d.h. Menschen, die Behandlung erbringen, wobei dies das Treffen von Entscheidungen oder die maßgebliche Mitwirkung an Entscheidungen über die Art der Behandlung mit einschließt,

- und einem Behandlungsteam (H&M)

$\mathbf{H\&M} := \mathbf{HCP} \cup \mathbf{M}$, mit $\mathbf{M} := \{\mathbf{M}_1, \dots, \mathbf{M}_{n_{\mathbf{M}}}\}$, $n_{\mathbf{M}} \geq 1$,

welches sowohl aus Menschen, die Behandlung erbringen, als auch aus intelligenten Maschinen besteht. \mathbf{M} schließt nur solche Maschinen ein, die ebenfalls Behandlung erbringen, wobei auch für \mathbf{M} gilt, dass dies das Treffen von Entscheidungen oder die maßgebliche Mitwirkung an Entscheidungen über die Art der Behandlung mit umfasst.

Entitäten in den Behandlungsteams HCP und H&M sollten natürlich hinreichend qualifiziert sein, um eine solche Behandlung kompetent erbringen zu können.



Die dritte Entitätsgruppe umfasst

(IR) Fragesteller (engl. interrogator) $\mathbf{IR} := \{\mathbf{IR}_1, \dots, \mathbf{IR}_{n_{\mathbf{IR}}}\}$, $n_{\mathbf{IR}} \geq 1$.

Analog zum Turing-Test sind dies Personen, die den Erfolg der Behandlung von Patienten ermitteln und analysieren. Auch diese Fragesteller sollten natürlich hinreichend qualifiziert sein, in diesem Fall um eine solche Bewertung kompetent vornehmen zu können.

Damit wie bei dem Turing-Test der bzw. die Fragesteller unvoreingenommen bleiben – dort konnte ein Fragesteller durch die Art der Kommunikation nicht unmittelbar erkennen, wer Mensch und wer Maschine ist – wäre es hier hilfreich, dass der bzw. die Fragesteller nicht wissen, ob die Behandlung von Patienten allein von Menschen oder gemeinsam von Mensch und Maschine durchgeführt wurde, einschließlich der dafür zu treffenden Entscheidungen. Auch bei der Unvoreingenommenheit von Patienten könnte es hilfreich sein, nicht zu wissen aus welchen Entitäten ihr Behandlungsteam besteht, um dadurch nicht in die eine oder andere auf den Behandlungserfolg einwirkende Richtung subjektiv beeinflusst zu werden.

3.3 Eine vorläufige, in Analogie zum Turing-Test formulierte Fassung des SYnENZ-Tests

3.3.1 Der Test

Es seien (PAT), (HCP), (H&M) und (IR) Entitäten wie 3.2 angegeben. (PAT) benötigen aufgrund eines Gesundheitsproblems (einer Erkrankung, ...) eine Behandlung (um eine Erkrankung zu heilen, zu lindern, einem Entstehen vorzubeugen, ...). (HCP) und (H&M) sind Behandlungsteams, welche die Kompetenz haben, diese Behandlung erbringen zu können. Bei (IR) liegt die Kompetenz vor, eine Bewertung des Erfolgs dieser Behandlung vorzunehmen.

Zudem sei

- (PAT) in zwei *Patientengruppen* aufgeteilt. In jeder Patientengruppe ist mindestens ein Patient. Eine der beiden Patientengruppen wird von (HCP) behandelt, die andere von (H&M).

In dieser Fassung des SYnENZ-Tests ermittelt (IR) die Behandlungsergebnisse all dieser Patienten und (IR) weiß, dass eine Patientengruppe von (HCP) und eine von (H&M) behandelt wurde. Um unvoreingenommen zu bleiben, weiß (IR) jedoch vor der Bewertung nicht, welche der beiden Patientengruppen von (HCP) und welche von (H&M) behandelt wurde.

Die zu Beginn gestellte Frage erweiterten Zusammenwirkens (engl.: extended collaboration) zwischen Menschen und Maschinen soll über zwei mögliche Aussagen (engl.: statements) beantwortet werden:

So Wenn (IR) keine Unterschiede in den Erfolgen der Behandlung von Patienten in den beiden Patientengruppen erkennen kann, dann kann das Zusammenwirken von Menschen und intelligenten Maschinen bei (H&M) nicht als erweitertes Zusammenwirken von Menschen und intelligenten Maschinen bezeichnet werden.

Wenn (IR) Unterschiede in den Erfolgen der Behandlung von Patienten in den beiden Patientengruppen erkennen kann (ohne, um unvoreingenommen zu bleiben, zu wissen, ob die Behandlung durch (HCP) oder durch (H&M) erfolgreicher war), muss anschließend festgestellt werden, welche Patientengruppe erfolgreicher behandelt wurde.



Wurde die erfolgreicher behandelte Patientengruppe durch (HCP) behandelt, dann kann das Zusammenwirken von Menschen und intelligenten Maschinen bei (H&M) ebenfalls nicht als erweitertes Zusammenwirken von Menschen und intelligenten Maschinen bezeichnet werden.

- Sec** Wurde die erfolgreicher behandelte Patientengruppe durch (H&M) behandelt, dann sollte dieses Zusammenwirken bei (H&M) als erweitertes Zusammenwirken von Menschen und intelligenten Maschinen bezeichnet werden.

3.3.2 Beispiel

Anhand eines fiktiven Beispiels soll verdeutlicht werden, wie ein solcher zwar erst vorläufiger, jedoch in dieser Variante dem Turing-Test ähnlicher SYnENZ-Test aussieht.

Die zu untersuchende Frage in diesem Beispiel: Bei Patienten, die an einer Erkrankung gelitten und die einen stationären Klinikaufenthalt gehabt haben, soll erreicht werden, dass diese nach Rückkehr in ihr normales Lebensumfeld wieder selbständig und selbstbestimmt leben und ggf. arbeiten können. Die hierfür zu erbringende Behandlung umfasst Maßnahmen der Gesundheitsversorgung, die diese Rückkehr in ein normales Lebensumfeld ermöglichen sollen und deren Erfolg gemessen werden kann. Zwei Behandlungsteams behandeln diese Patienten: ein interdisziplinäres Team mit Ärzten, Pflegekräften und Physiotherapeuten sowie ein Team, das ebenfalls aus diesen Fachkräften in Gesundheitsberufen besteht, indem jedoch auch eine intelligente, auf diese Behandlung spezialisierte Maschine beteiligt ist.

Der Verfasser hat sich bei diesem fiktiven Beispiel an zwei konkret durchgeführte Studien angelehnt: der GAL-NATARS-Studie ([13], [14], [15]) und der AGT-Reha-WK-Studie ([16], [17], [18]). Beide Studien sind in [19] als Beispiele einführend (und, so hofft der Verfasser, prägnant) beschrieben. In diesen Studien ging es um Sturz-Patienten (GAL-NATARS-Studie) bzw. um Patienten mit Schulterbeschwerden (AGT-Reha-WK-Studie), bei denen untersucht wurde, inwieweit nach deren stationären Klinikaufenthalt Leben und ggf. Arbeiten im normalen Lebens- bzw. Arbeitsumfeld wieder möglich war.

Es sei eine dem Turing-Test vergleichbare – wohlgernekt fiktive – Untersuchung mittels des vorläufigen SYnENZ-Tests durchgeführt worden. IR wurde berichtet, dass die Behandlungsteams HCP und H&M jeweils 10 Patienten behandelt hatten. Ohne dass IR schon wusste, welches Behandlungsteam welche Patientengruppe behandelt hatte, wurden IR die Ergebnisse vorgelegt. In der ersten Gruppe wurden 6 Patienten erfolgreich behandelt, in der zweiten Gruppe 5. Aufgrund der geringen Differenz entschied IR, dass Aussage $S_{\mathcal{A}}$ zu gelten hat, also dass das Zusammenwirken von Menschen und intelligenten Maschinen bei (H&M) nicht als erweitertes Zusammenwirken von Menschen und intelligenten Maschinen bezeichnet werden. Anschließend erfuhr IR, dass die erste Patientengruppe von H&M und die zweite von HCP behandelt wurde.

3.3.3 Vor- und Nachteile

Diese vorläufige Fassung des SYnENZ-Tests soll die Analogie des SYnENZ-Tests zum Turing-Test verdeutlichen. Unterschiede, zumindest wesentliche, gibt es bei der Frage, die transportiert aufgrund des Tests beantwortet werden soll – F_{SYnENZ} anstelle von F_{Turing} . Zudem wird bei den zu befragenden Entitäten weiter differenziert – im Turing-Test sind es Menschen und Maschinen; im SYnENZ-Test sind es Patienten, die eine Behandlung benötigen sowie Fachkräfte in Gesundheitsberufen oder Fachkräfte in Gesundheitsberufen gemeinsam mit Maschinen, die diese Behandlung erbringen.



Wie der Turing-Test hat dieser vorläufige SYnENZ-Test den Vorteil, dass er leicht zu verstehen ist. Zudem dürfte es auch hier plausibel sein, dass die Frage des erweiterten Zusammenwirkens zwischen Menschen und intelligenten Maschinen dann wohl mit ja beantwortet werden muss, wenn in dem SYnENZ-Test die Aussage S_{ec} zutrifft. Nicht zuletzt ist es gut vorstellbar ist, Untersuchungen mit solchen SYnENZ-Tests durchführen zu können.

Aber es gibt auch einige Schwächen. So ist z.B. nicht klar

- wie Patienten für diesen Test ausgewählt wurden,
- wie viele Patienten aufgrund welcher Überlegungen dazu in die Untersuchung eingeschlossen wurden,
- wie die Zuordnung von Patienten zu den Behandlungsgruppen von HCP bzw. von H&M erfolgte und
- wie die Aussage, ob ein erweitertes Zusammenwirken besteht (S_{ec}) oder nicht (S_{\neq}), auf Basis der Behandlungsergebnisse durch (IR) abgeleitet wurde.

Die nachfolgend vorgeschlagene erweiterte Fassung soll diese Schwächen beseitigen.

3.4 Die vollständige Fassung des SYnENZ-Tests

3.4.1 Hinweise zur Methodik

Für diese vollständige Version des SYnENZ-Tests werden wir 'klassische' Ansätze der Versuchsplanung und des statistischen Testens nutzen. Das Fachgebiet, das sich im Anwendungsgebiet Medizin und Gesundheitsversorgung mit dieser Methodik beschäftigt, ist die Medizinische Biometrie bzw. Medizinische Statistik. Einführende Lehrbücher sind z.B. [20], [21] und [22]. Für Interessierte an der Entwicklung dieses Gebietes in Deutschland sei auf [23] und [24] hingewiesen.

In der statistischen Testtheorie ziehen wir eine Stichprobe aus einer Grundgesamtheit, z.B. von Patienten, die an einer bestimmten Erkrankung leiden. Diese Patienten können, z.B., mit zwei verschiedenen Therapien behandelt werden. Beim Testen dieser Stichprobe unterscheiden wir zwischen einer Nullhypothese (oft als H_0 bezeichnet), dass kein Unterschied zwischen diesen beiden Behandlungen gefunden werden kann, und einer Alternativhypothese (oft als H_1 bezeichnet), dass wir annehmen müssen, dass Unterschiede zwischen diesen beiden Therapien existieren.

Auf weitere Details z.B. zu Fehlerarten (Fehler 1. und 2. Art) und deren Wahrscheinlichkeiten (meist mit a und b bezeichnet), zur Stichprobenumfangsberechnung (z.B. auf Basis von Erfolgsraten der Therapien) oder zu Studientypen (Beobachtungsstudien, Interventionsstudien, ...) soll hier nicht eingegangen werden. Es ist jedoch festzuhalten, dass randomisierte Studien, in denen Patienten streng zufällig zu Therapien zugeteilt werden, weltweit zum 'Goldstandard' bei Interventionsstudien geworden sind ([24], S. 6), und dass eine Verblindung – einfache Verblindung: Patienten wissen nicht, welche Behandlung sie erhalten haben, doppelte Verblindung: Patienten und Behandelnde wissen nicht, welche Behandlung die Patienten erhalten haben – hilfreich ist, um Behandlungserfolge unvoreingenommen bewerten zu können ([20], S. 32-33).

Zu beachten ist, dass in der statistischen Testtheorie Entscheidungen über H_0 und H_1 asymmetrisch sind: Entweder kann H_0 nicht verworfen werden oder H_1 muss angenommen werden. Dies ist ein Unterschied zu den symmetrischen Entscheidungen durch die Aussagen S_{\neq} und S_{ec} in der vorläufigen Version.

Zudem sei darauf hingewiesen, dass bei Studien mit Patienten strenge ethische und rechtliche Bestimmungen eingehalten werden müssen. Dies gilt insbesondere für die Planung, Durchführung und Auswertung klinischer Studien, die gerade bei Therapiestudien ja Experimente am Menschen sind und für die die Deklaration von Helsinki des Weltärztebundes ([25]) eine wichtige Grundlage bildet. Diese Vorschriften müssen berücksichtigt werden, werden aber auch hier nicht weiter besprochen.

3.4.2 Der Test

Es seien wieder (PAT), (HCP), (H&M) und (IR) Entitäten wie 3.2 angegeben. (PAT) benötigen aufgrund eines Gesundheitsproblems (einer Erkrankung, ...) eine Behandlung (um eine Erkrankung zu heilen, zu lindern, einem Entstehen vorzubeugen, ...). (HCP) und (H&M)



sind Behandlungsteams, welche die Kompetenz haben, diese Behandlung erbringen zu können. Bei (IR) liegt die Kompetenz vor, eine Bewertung des Erfolgs dieser Behandlung vorzunehmen.

Zudem sei erneut

- (PAT) in zwei Patientengruppen aufgeteilt. Eine Patientengruppe wird von (HCP) behandelt, die andere von (H&M).

Wie schon in der vorläufigen Fassung des SYnENZ-Tests ermittelt (IR) die Behandlungsergebnisse all dieser Patienten und (IR) weiß, dass eine Patientengruppe von (HCP) und eine von (H&M) behandelt wurde. Um unvoreingenommen zu bleiben, weiß (IR) jedoch erneut nicht, welche der beiden Patientengruppen von (HCP) und welche von (H&M) behandelt wurde.

Und wie in der einfachen Fassung soll die zu Beginn gestellte Frage erweiterten Zusammenwirkens zwischen Menschen und Maschinen in transponierter Form über den SYnENZ-Test beantwortet werden.

In der erweiterten Fassung des SYnENZ-Tests verwenden wir vergleichende Interventionsstudien als Studientyp, die bevorzugt als randomisierte Studien geplant und durchgeführt werden sollten: Jeder Patient $PAT_1, \dots, PAT_{n_{PAT}}$ wird streng zufällig einer der beiden Patientengruppen zugeteilt, von denen die eine durch HCP und die andere durch H&M behandelt wird.

In einer solchen randomisierten Studie und auf der Grundlage eines geeigneten statistischen Tests testen wir die Hypothesen

$H-\text{ec}_0$ Es gibt keinen Unterschied bei dem Behandlungserfolg durch die beiden Behandlungsteams und damit auch kein erweitertes Zusammenwirken zwischen Menschen und intelligenten Maschinen (also die Nullhypothese, ' H_0 ').

$H\text{ec}_?$ Es gibt einen Unterschied bei dem Behandlungserfolg durch die beiden Behandlungsteams (also die Alternativhypothese, ' H_1 '). Hier muss dann unterschieden werden:

$H-\text{ec}_1$ Ist (HCP) das erfolgreichere Behandlungsteam, dann kann das Zusammenwirken bei (H&M) nicht als erweitertes Zusammenwirken von Menschen und intelligenten Maschinen bezeichnet werden.

$H\text{ec}$ Ist (H&M) das erfolgreichere Behandlungsteam, dann sollte dieses Zusammenwirken bei (H&M) als erweitertes Zusammenwirken von Menschen und intelligenten Maschinen bezeichnet werden.

In anderen Worten: Falls (IR) nach der Durchführung einer randomisierten Studie zu einer bestimmten Behandlung von Patienten und nach der Analyse der Behandlungsergebnisse mit Hilfe dieses statistischen Tests

- keine Unterschiede in den Erfolgen der Behandlung von Patienten in den beiden Patientengruppen erkennen kann, dann kann die Nullhypothese, dass es kein erweitertes



Zusammenwirken zwischen Menschen und intelligenten Maschinen gibt, nicht verworfen werden.

Wenn aber (IR)

- Unterschiede in den Erfolgen der Behandlung von Patienten in den beiden Patientengruppen erkennen kann und damit die Alternativhypothese angenommen werden muss, dann gibt es zwei Fälle zu unterscheiden:
 - Ist (HCP) das erfolgreichere Behandlungsteam, dann gibt es kein,
 - ist jedoch (H&M) das erfolgreichere Behandlungsteam, dann gibt es ein erweitertes Zusammenwirken von Menschen und intelligenten Maschinen.

3.4.3 Beispiel

Anhand eines fiktiven Beispiels soll wieder verdeutlicht werden, wie ein solcher (nun vollständiger) SYnENZ-Test aussehen kann.

Die zu untersuchende Frage sei dieselbe wie in dem ersten Beispiel. Bei Patienten, die an einer Erkrankung gelitten und die einen stationären Klinikaufenthalt gehabt haben, soll erreicht werden, dass diese nach Rückkehr in ihr normales Lebensumfeld wieder selbstständig und selbstbestimmt leben und ggf. arbeiten können. Die hierfür zu erbringende Behandlung umfasst Maßnahmen der Gesundheitsversorgung, die diese Rückkehr in ein normales Lebensumfeld ermöglichen sollen und deren Erfolg gemessen werden kann. Zwei Behandlungsteams behandeln diese Patienten: ein interdisziplinäres Team mit Ärzten, Pflegekräften und Physiotherapeuten sowie ein Team, das ebenfalls aus diesen Fachkräften in Gesundheitsberufen besteht, indem jedoch auch eine intelligente, auf diese Behandlung spezialisierte Maschine beteiligt ist.

Wie in 3.3.2 weiter ausgeführt hat sich der Verfasser auch bei diesem fiktiven Beispiel wieder an den beiden konkret durchgeführten Studien angelehnt: der GAL-NATARS-Studie und der AGT-Reha-WK-Studie (einführende Beschreibung in [19]).

Zunächst muss ein statistisches Modell erstellt werden.

Nehmen wir an, dass das Behandlungsergebnis dichotom ist und dass das Ergebnis entweder erfolgreich (+) oder nicht erfolgreich (-) ist.

Es bezeichne $P_{ij}, i \in \{HCP, H\&M\}, j \in \{+, -\}$ die Wahrscheinlichkeit, dass die Behandlung einer Person durch das Behandlungsteam i den Behandlungserfolg j erbracht hat.

Unsere Hypothesen zum erweiterten Zusammenwirken von Menschen und Maschinen lauten:

$$\mathbf{H-ec}_0 \quad P(j \mid HCP) = P(j \mid H\&M) \text{ oder anders formuliert } P_{ij} = P_i \cdot P_j \quad \text{für alle } ij.$$

Es gibt keine Erfolgsunterschiede in der Behandlung durch die beiden Behandlungsteams und damit auch kein erweitertes Zusammenwirken zwischen Menschen und intelligenten Maschinen (also die Nullhypothese, ' \mathbf{H}_0 ').

Falls (IR) nicht zwischen den Behandlungsergebnissen der beiden Gruppen unterscheiden kann oder wenn die Gruppe mit dem besseren Ergebnis von (HCP)

behandelt wurde, kann die Zusammenarbeit von (H&M) nicht als erweitertes Zusammenwirken bezeichnet werden;

Hec₂ $P(j | HCP) \neq P(j | H\&M)$ oder anders formuliert $P_{ij} \neq P_i \cdot P_j$ für mindestens ein ij .

Es gibt Erfolgsunterschiede in der Behandlung durch die beiden Behandlungsteams (also die Alternativhypothese, 'H₁'). Hier muss dann unterschieden werden:

H-ec₁ $P(+ | HCP) > P(+ | H\&M)$; ist (HCP) das erfolgreichere Behandlungsteam, dann kann das Zusammenwirken bei (H&M) nicht als erweitertes Zusammenwirken von Menschen und intelligenten Maschinen bezeichnet werden.

Hec $P(+ | H\&M) > P(+ | HCP)$; ist (H&M) das erfolgreichere Behandlungsteam, dann sollte dieses Zusammenwirken bei (H&M) als erweitertes Zusammenwirken von Menschen und intelligenten Maschinen bezeichnet werden.

Als statistischer Test kann hier der 'klassische' χ^2 -Test nach Pearson ([26]) genutzt werden (vgl. z.B. [20], S. 287-290). Die Teststatistik lautet:

$$T := \sum_{i,j} (n_{ij} \cdot \hat{E}(n_{ij})^2 / \hat{E}(n_{ij}) \quad \text{mit} \quad \hat{E}(n_{ij}) := n_i \cdot n_j / n_{..} \quad .$$

Wir wollen in diesem Beispiel weiterhin annehmen, dass die Behandlung dieser Patienten durch ein interdisziplinäres Team mit Ärzten, Pflegekräften und Physiotherapeuten schon länger eingeführt ist und dass die Erfolgsrate dieser Behandlung bei 70% liegt. Die Behandlung durch Behandlungsteams, in denen zusätzlich zu solchen Fachkräften in Gesundheitsberufen auch eine intelligente, auf diese Behandlung spezialisierte Maschine beteiligt ist, soll nun neu erprobt werden. Um diese neue Art der Behandlung einzuführen wären erhebliche Aufwände nötig. Es wird festgelegt, dass die Erfolgsrate um 10% höher als bei der bereits eingeführten Behandlung liegen und damit mindestens 80% betragen soll.

Auf Basis dieser Festlegung und nach der weiteren Festlegung, dass der Fehler 1. Art bei $\alpha := 5\%$ sowie der Fehler 2. Art bei $\beta := 10\%$ liegen soll, kann der notwendige Stichprobenumfang bestimmt werden: Er beträgt hier 176 Patienten.

Informationen zur Bestimmung eines passenden Stichprobenumfangs finden sich z.B. in [20], S. 444-452. Diese Berechnung wurde mit Hilfe von G*Power ([27], [28]) ermittelt. Die Umrechnung der anstelle der Erfolgsraten benötigten Effektstärke, hier 0,245, erfolgte mit Hilfe von [29].

Es sei eine – wohlgermerkt wieder fiktive – Studie durchgeführt worden, und zwar als randomisierte Doppelblindstudie. Die 176 Patienten werden streng zufällig auf die beiden Behandlungsteams HCP und H&M aufgeteilt. Doppelblind bedeutet in diesem Fall, dass weder Patienten wissen, ob deren Behandlungsteam neben Fachkräften in Gesundheitsberufen auch aus intelligenten Maschinen besteht und dass die Fragesteller ebenfalls nicht wissen, welche Patientengruppen durch welche Behandlungsteams behandelt wurden.



Die Studie brachte folgendes Ergebnis

Behandlungs- team	Behandlungserfolg		Σ
	+	-	
H&M	63 (72%)	25 (28%)	88
HCP	59 (67%)	29 (33%)	88
Σ	122 (69%)	54 (67%)	176 (31%)

Es ergibt sich als Teststatistik der Wert $T = 0,43$ (kritische Schranke: 3,84, p-Wert: 0,51). Insofern kann **Hec?**, d.h. die Nullhypothese, dass kein erweitertes Zusammenwirken zwischen Menschen und intelligenten Maschinen besteht, nicht verworfen werden.

3.4.4 Vor- und Nachteile

Wie in der vorläufigen Fassung hat auch diese vollständige Fassung des SYNENZ-Tests den Vorteil, dass die Umsetzung der Frage erweiterten Zusammenwirkens sinnvoll erscheint und dass es gut vorstellbar ist, Studien mittels solcher SYNENZ-Tests durchführen zu können.

Der Test sollte zumindest in seinem Testaufbau und in seinen Hypothesen immer noch leicht zu verstehen sein, obwohl mehr methodische Kenntnisse in der Versuchsplanung und im statistischen Testen notwendig sind, um Details zu beurteilen zu können.

Auch die hier explizit erwähnte Differenzierung der beiden Fälle nicht vorhandenen Zusammenwirkens – im Fall der Nullhypothese kann diese nicht verworfen werden, im Fall der Alternativhypothese wird entschieden, dass dieses Zusammenwirken nicht vorliegt – erscheint komplizierter. Hier ist zu ergänzen, dass der Test in dem Beispiel als zweiseitiger Test ausformuliert wurde. Hätte man einen einseitigen Text gewählt (was durchaus Sinn macht), dann gäbe es weiterhin nur eine Nullhypothese **H-ec**, die nicht verworfen werden könnte oder eine Alternativhypothese **Hec**, die anzunehmen wäre.

Und der durchaus wesentliche Vorteil, die Schwächen der vorläufigen Fassung des Tests zu beheben, hat gleichzeitig den Nachteil, dass die Planung, Durchführung und Auswertung solcher Experimente aufwändiger wird.

Ein weiterer Vorteil ist, dass wir die gut ausgearbeitete Methodik zur Versuchsplanung und zum statistischen Testen nutzen können. Im Anwendungsgebiet Medizin und Gesundheitsversorgung ist diese Methodik seit Jahrzehnten erfolgreich etabliert ([24]).

4 Diskussion

4.1 Weitere Hinweise zu dieser Ausarbeitung

Diese Ausarbeitung beschreibt methodische Überlegungen in einem frühen Stadium. Durchaus wünschenswert wäre gewesen, wenn zumindest schon eine konkrete Studie mit dem SYNENZ-Test durchgeführt worden wäre. Andererseits sollten diese methodischen Überlegungen aber auch frühzeitig kommuniziert werden, da sie evtl. auch für andere Personen hilfreich sind, um Fragen des erweiterten Zusammenwirkens empirisch überprüfen zu wollen.



Zwei Punkte mit Bezug zu Alan Turings eingangs zitierter Arbeit: Mit dem SYnENZ-Test lässt sich (wie gesagt aus Sicht des Verfassers) empirisch nachweisen, ob erweitertes Zusammenwirken zwischen Menschen und intelligenten Maschinen besteht. Warum dies so ist und auf welchen Prinzipien oder Methoden dies beruht, erklärt der Test nicht. In [7] versucht Alan Turing, nach Vorstellung seines Tests, zu überlegen, wie solche Prinzipien oder Methoden besser verstanden werden können. Interessant ist zudem, warum er bei seinem Test nicht auch die damals schon einigermaßen etablierte Versuchsplanung, verbunden mit der Anwendung statistischer Testtheorie angewandt hat – der hier genutzte χ^2 -Test nach Pearson wurde z.B. fast 50 Jahre früher veröffentlicht. Eine Möglichkeit könnte sein, dass die Anwendungsgebiete solcher empirischen Ansätze, u.a. damals die Landwirtschaft und die vor allem der angewandten Statistik zuordenbaren methodischen Überlegungen, bei ihm und bei Arbeiten in seinem Themenbereich nicht hinreichend bekannt waren.

Der SYnENZ-Test liefert eine Vorgehensweise, mit der erweitertes Zusammenwirken festgestellt werden kann. Es gibt allerdings Ermessensspielräume bei dem Testaufbau: Wann können Maschinen als „intelligente Maschinen“ bezeichnet werden bzw. wann sind sie „funktional umfassend“ und treffen „komplexe Entscheidungen“ (vgl. 1.2)? Zudem hängt hier die Aussage des erweiterten Zusammenwirkens von dem Erfolg bei dem (gemeinsamen) Lösen einer Aufgabe ab. Dieses Ergebnis-orientierte Denken ist typisch für die Medizin und Gesundheitsversorgung, insbesondere bei der Behandlung von Patienten, und hat sich dort bewährt. Dies muss aber nicht überall gelten.

In der Therapieforschung und dort bei klinischen Prüfungen werden bei der Evaluation üblicherweise vier Phasen unterschieden: Erst nachdem die Phasen I (Pharmakokinetik, Dosis-Wirkungsbeziehung) und II (Verträglichkeit, prinzipielle Wirksamkeit) erfolgreich durchlaufen wurden, darf in einer Phase III mittels kontrollierter klinischer Studien ein Wirksamkeitsnachweis an einer größeren Zahl von Patienten untersucht werden. Es erfolgt dann, nach Zulassung, in einer Phase IV eine weitere Überwachung des Therapieerfolgs und möglicher unerwünschter Wirkungen (vgl. z.B. [20], S. 51). Der hier vorgeschlagene SYnENZ-Test untersucht ganz allgemein die Frage erweiterten Zusammenwirkens. In diesem Schema lässt er sich insbesondere in Phase III einordnen. In [30] beschreibt David Spiegelhalter in einer thematisch verwandten Arbeit Evaluationsansätze für intelligente Maschinen (wohl ohne die Ausrichtung auf erweitertes Zusammenwirken), in der dieses 4-Phasen-Schema explizit aufgriffen und dessen Anwendung empfohlen wird.

4.2 Erweiterbarkeit auf andere Szenarien erweiterten Zusammenwirkens

Wie in 3.1 geschrieben, beschränkt sich in dieser Ausarbeitung die Beschreibung des SYnENZ-Tests auf das erweiterte Zusammenwirken von Menschen und Maschinen bei dem Anwendungsfall der Behandlung von Patienten. Eine Erweiterung

- auf andere Anwendungsfälle im Anwendungsgebiet Medizin und Gesundheitsversorgung und auf Anwendungsfälle in anderen Anwendungsgebieten dürfte dann geben, wenn die jeweiligen Anwendungsfälle einen ähnlichen Aufbau haben: Es gibt eine *Aufgabe*, die von zwei *Teams*, zum einen einer Gruppe von Menschen und zum anderen von einer Gruppe von Menschen und intelligenten Maschinen bearbeitet wird, die anhand von Erfolgskriterien beurteilt werden und die wiederholt durchgeführt werden kann. Eine Erweiterung
- auf andere Entitäten sollte ohne Einschränkungen möglich sein.

Andere, hier nicht behandelte und gut möglich erscheinende Erweiterungen sind die Erweiterung auf mehr als zwei Gruppen von Bearbeitern, die in der Versuchsplanung anstelle



von Studienplänen mit zwei Stichproben auf Studienpläne mit mehr als zwei Stichproben abgebildet werden können und die bei Aufstellung des statistischen Modells und der Hypothesen entsprechend berücksichtigt werden müssen.

Interessant ist natürlich auch die Frage, wenn eine Gruppe von Bearbeitern aus der Menge M besteht, dies besonders dann, wenn das Ergebnis eines SYnENZ-Tests beispielsweise bei der Behandlung von Patienten eine Überlegenheit von M gegen über HCP oder H&M ergeben hätte.

Der SYnENZ-Test dürfte auf zahlreiche Anwendungsszenarien anwenden lassen. Dazu zählen Fragen der gemeinsamen Behandlung von Patienten durch Menschen und intelligente Maschinen, sei es in der Prävention, der Diagnostik, der Therapie, der Pflege oder in Kombinationen davon (vgl. z.B. [4] und die dortigen Literaturverweise). Hingewiesen sei hier auf die Möglichkeit neuer Versorgungsformen durch solche gemeinsamen Behandlungen, beispielsweise durch entsprechend ausgestattete Wohnungen, die sich als intelligente Maschinen an der Betreuung von Personen mit Gesundheitsrisiken (z.B. Gebrechlichkeit, erhöhtes Sturzrisiko) beteiligen und bei der die Wohnung sozusagen als maschineller Diener an der Gesundheitsversorgung von Personen beteiligt ist (weitere Details ebenfalls in [4], vgl. auch [6], [15]).

4.3 Der SYnENZ-Test und die Leitfragen der SYnENZ-Kommission

Welchen Beitrag könnte die Nutzung des SYnENZ-Tests für die Beantwortung in 1.1 beschriebenen Leitfragen der SYnENZ-Kommission leisten? Die fünf Fragen lauteten:

- 1 Wie wird unser Zusammenleben in Zeiten zunehmender Digitalisierung in Zukunft aussehen?
- 2 Welche Synergien ergeben sich durch das nun mögliche erweiterte Zusammenwirken von Menschen, Tieren und Pflanzen einerseits und von Maschinen andererseits, d.h. von natürlicher und künstlicher Intelligenz?
- 3 Können wir zwischen bloß zeitgemäßen und angemessenen Formen des Zusammenwirkens unterscheiden?
- 4 Und können wir Empfehlungen geben, um angemessene Formen zu erreichen und um zwar zeitgemäße, aber problematische Formen zu vermeiden?
- 5 Lässt sich ein Grad des Zusammenwirkens bestimmen?

Die Nutzung des SYnENZ-Tests wird – leider – keine Antworten auf diese Fragen liefern. Der Test mag jedoch dazu beitragen, Synergien, die sich aus einem möglichen erweiterten Zusammenwirkens ergeben, besser zu untersuchen und nachvollziehbarere Aussagen darüber zu machen (Beitrag zur Frage 2) sowie den Grad des Zusammenwirkens zu bestimmen (Beitrag zur Frage 5).

4.4 Die Frage des erweiterten Zusammenwirkens transponiert auf den SYnENZ-Test

Mit dem SYnENZ-Test und sozusagen in Analogie zum Turing-Test kann damit die eingangs gestellte Frage erweiterten Zusammenwirkens jetzt transponiert werden auf eine andere Frage, die über diesen SYnENZ-Test – zumindest aus Sicht des Verfassers – empirisch ermittelt und dadurch beantwortet werden kann:

FSynENZ (transponiert): Muss H_{ec} nach einem SYnENZ-Test angenommen werden?
Falls ja, dann kann wohl auch F_{SYnENZ} : Kann erweitertes Zusammenwirken zwischen Menschen und intelligenten Maschinen festgestellt werden? mit ja beantwortet werden.



5 Ein Aufruf zur Diskussion und zur Erprobung

Der Verfasser wäre dankbar für eine breite kritische Diskussion der vorgestellten Überlegungen. Noch dankbarer wäre er, wenn nicht nur er, sondern auch andere den SYnENZ-Test nutzen würden, um Fragen des erweiterten Zusammenwirkens zwischen Menschen und intelligenten Maschinen zu untersuchen. Erfahrungsgemäß helfen empirische Untersuchungen der Versachlichung von Themen, reduzieren spekulative Debatten und konkretisieren Aktivitäten in Forschung und Anwendung. Dies hilft – und hier soll nochmals der in 1.1. genannte Gegenstand der Arbeit der SYnENZ-Kommission aufgegriffen werden – sowohl die technischen als auch die ethischen und die rechtlichen Herausforderungen des Zusammenwirkens lebender und nicht lebender Entitäten, die im Zeitalter der Digitalisierung möglich geworden sind, gezielter zu bearbeiten.

Danksagung

Der Verfasser ist vielen Personen Dank schuldig, die zum Entstehen dieser Ausarbeitung beigetragen haben. Zu nennen sind die Mitglieder der SYnENZ-Kommission der BWG, dort u.a. Otto Luchterhandt (der nach der Bedeutung des Begriffs des erweiterten Zusammenwirkens nachgefragt und dadurch diese Ausarbeitung motiviert hatte), Jochen Steil (dem Organisator des ersten SYnENZ-Zirkels, der den Verfasser u.a. gebeten hatte, über die Turing-Arbeit von 1948 zu referieren und der sich damit nochmals intensiv mit dieser Arbeit beschäftigen dürfte; zudem ermöglichte er dem Verfasser eine erste Vorstellung und Diskussion seiner Überlegungen), Stefan Heuser (dem Organisator des zweiten SYnENZ-Zirkels, dass der Verfasser das Eingangsreferat zu diesem Zirkel, der über Grade des Zusammenwirkens ging, halten und dadurch diese Ausarbeitung vorstellen durfte) sowie Otto Richter (der diese Fassung für das BWG-Jahrbuch erbeten hatte und damit genügend Termindruck für die Fertigstellung herstellte). Dank geht auch an meine Medizininformatik-Kollegen Klaus-Hendrik Wolf und Michael Marscholke in der Kommission für nicht mehr zu zählende Diskussionen dazu. Wertvolle inhaltliche und Literaturhinweise erhielt der Verfasser auch durch Bianca Steiner, Martin Schumacher, Olaf Gefeller und Rudolf Kruse. Nicht zuletzt wurde er bei der Literaturlarbeit unterstützt von Nadine Maxrath und Ute Zeisberg.

Begriffe

Begriff	Abschnitt in dem der Begriff eingeführt wurde	Erläuterung
<i>Aufgabe</i>	4.2	Im SYnENZ-Test: Verallgemeinerung von Behandlung (weitere Details: siehe dort).
<i>Behandelnde</i>	3.1 3.2	erbringen Behandlung, d.h. Maßnahmen der Gesundheitsversorgung, wobei dies das Treffen von Entscheidungen oder die maßgebliche Mitwirkung an Entscheidungen über die Art der Behandlung mit einschließt,
<i>Behandlung ...</i>	3.1	... von Patienten. Maßnahmen der Gesundheitsversorgung, die in der Regel durch Fachkräfte in Gesundheitsberufen (Ärztinnen/Ärzte, Pflegekräfte, ...) für Patienten erbracht werden. Diese Maßnahmen der Gesundheitsversorgung können Maßnahmen der Prävention, der Diagnostik, der Therapie und der Pflege umfassen. Die gemachte Einschränkung „in der Regel“



- bezieht sich hier darauf, dass die Behandlung von Patienten auch von anderen intelligenten Maschinen erbracht werden könnte.
- 4.2 Verallgemeinerung auf weitere Szenarien erweiterten Zusammenwirkens: Aufgabe, die zum einen von einer Gruppe von Menschen und zum anderen von einer Gruppe von Menschen und intelligenten Maschinen bearbeitet wird und anhand von Erfolgskriterien beurteilt werden wiederholt durchgeführt werden kann.
- Behandlungsteam* 3.2 Im SYnENZ-Test gibt es zwei Teams mit Behandelnden. Das eine Team besteht aus Fachkräften in Gesundheitsberufen und das andere aus solchen Fachkräften gemeinsam mit intelligenten Maschinen.
- 4.2 Verallgemeinerung auf weitere Szenarien erweiterten Zusammenwirkens: Im SYnENZ-Test gibt es zwei Teams, die zum einen aus einer Gruppe von Menschen und zum anderen aus einer Gruppe von Menschen und intelligenten Maschinen bestehen. Die beiden Teams haben die gleiche Aufgabe zu bearbeiten.
- Entitäten ...* 1.1 ..., mit denen sich die BWG-Kommission Synergie und Intelligenz befasst, sind lebende Entitäten (Menschen, aber auch z.B. Tiere und Pflanzen) sowie nicht lebende Entitäten (Maschinen, ...).
- Entitätsgruppen ...* 3.2 ... im SYnENZ-Test sind Patienten, Behandelnde, Befrager.
- intelligente Maschinen* 1.2 Maschinen, deren Funktionalität so umfangreich ist, dass diese Maschinen komplexe Entscheidungen, wie beispielsweise in der Diagnostik oder wie bei Therapien, treffen können oder zumindest maßgeblich an solchen Entscheidungen mit beteiligt sind. Solche Maschinen sind üblicherweise realisiert als auf Computern implementierte Programme. Ab Abschnitt 3.1 werden diese auch verkürzt als Maschinen bezeichnet, da ab diesem Abschnitt in dieser Ausarbeitung bei Maschinen immer intelligente Maschinen gemeint sind.
- Maschinen* 3.1 Ab Abschnitt 3.1 werden intelligente Maschinen auch verkürzt als Maschinen bezeichnet, da ab diesem Abschnitt in dieser Ausarbeitung bei Maschinen immer intelligente Maschinen gemeint sind.
- Patienten* 3.1 Menschen, die Behandlung, d.h. Maßnahmen der Gesundheitsversorgung, benötigen.
- Patientengruppe* 3.3 Im SYnENZ-Test gibt es zwei Patientengruppen. Die Patienten der einen Gruppe werden von einem Behandlungsteam, das aus Fachkräften in Gesundheitsberufen besteht, behandelt. Die Patienten der anderen Gruppe werden von einem Behandlungsteam, das Fachkräften in Gesundheitsberufen und aus intelligenten Maschinen besteht, behandelt.
- Team* 4.2 Im SYnENZ-Test: Verallgemeinerung von Behandlungsteam (weitere Details: siehe dort).



Literatur

- [1] Kommission Synergie und Intelligenz: technische, ethische und rechtliche Herausforderungen des Zusammenwirkens lebender und nicht lebender Entitäten im Zeitalter der Digitalisierung (SYnENZ) der Braunschweigischen Wissenschaftlichen Gesellschaft.
<http://bwg-nds.de/kommissionen/kommission-synenz/>. Zugriffen: 3. Februar 2021.
- [2] Braunschweigische Wissenschaftliche Gesellschaft (BWG). <http://bwg-nds.de>. Zugriffen: 3. Februar 2021.
- [3] Jahresberichte der BWG-Kommission Synergie und Intelligenz: technische, ethische und rechtliche Herausforderungen des Zusammenwirkens lebender und nicht lebender Entitäten im Zeitalter der Digitalisierung (SYnENZ). Die Berichte erscheinen in den Jahrbüchern der Braunschweigischen Wissenschaftlichen Gesellschaft.
Online bei <http://bwg-nds.de/veroeffentlichungen-jahrbuch-und-abhandlungen/>. Zugriffen: 3. Februar 2021.
- [4] Haux R. Über das Zusammenwirken von Menschen und 'intelligenten' Maschinen in der Gesundheitsversorgung. Was kann und was soll Informatik leisten? In: Manzeschke A, Niederlag W (Hrsg.). Ethische Perspektiven auf Biomedizinische Technologie, 182-97. Berlin: de Gruyter; 2020.
- [5] Schilling M, Kopp S, Wachsmuth S, Wrede B, Ritter H, Brox T, Nebel B, Burgard W. Towards a multidimensional perspective on shared autonomy. Proc. AAAI Fall Symposium Series, 338-44. Stanford; 2016.
- [6] Steil J, Finas D, Beck S, Manzeschke A, Haux R. Robotic systems in operating theaters: New forms of team-machine interaction in health care. Methods Inf Med. 2019; 58: e14-25.
- [7] Turing AM. Computing machinery and intelligence. Mind. 1950; 59: 433-60.
- [8] DeepL GmbH, www.deepl.com. Zugriffen: 3. Februar 2021.
- [9] Copeland BJ. The Turing Test. Minds and Machines. 2000; 10: 519-39.
- [10] Saygin A, Cicekli I, Akman V. Turing test: 50 years later. Minds and Machines. 2000; 10: 463-518.
- [11] van Bommel JH. Where were you? Reflections on science, philosophy and religion. 2015.
- [12] Gahl K. Begegnung und Verantwortung. Beiträge zu einer ärztlichen Menschenkunde. Karl Alber: Freiburg; 2019.
- [13] Marschollek M, Becker M, Bauer JM, Bente P, Dasenbrock L, Elbers K, Hein A, Kolb G, Künemund H, Lammel-Polchau C, Meis M, Meyer Zu Schwabedissen H, Remmers H, Schulze M, Steen EE, Thoben W, Wang J, Wolf KH, Haux R. Multimodal activity monitoring for home rehabilitation of geriatric fracture patients – feasibility and acceptance of sensor systems in the GAL-NATARS study. Inform Health Soc Care. 2014; 39: 262-71.
- [14] Wang J, Bauer J, Becker M, Bente P, Dasenbrock L, Elbers K, Hein A, Kohlmann M, Kolb G, Lammel-Polchau C, Marschollek M, Meis M, Remmers H, zu Schwabedissen HM, Schulze M, Steen EE, Haux R, Wolf KH. A novel approach for discovering human behavior patterns using unsupervised methods. Z Gerontol Geriatr. 2014; 47: 648–60.
- [15] Haux R, Hein A, Kolb G, Künemund H, Eichelberg M, Appell JE, Appellath HJ, Bartsch C, Bauer JM, Becker M, Bente P, Bitzer J, Boll S, Büsching F, Dasenbrock L, Deparade R, Depner D, Elbers K, Fachinger U, Felber J, Feldwieser F, Forberg A, Gietzelt M, Goetze S, Gövercin M, Helmer A, Herzke T, Hesselmann T, Heuten W, Huber R, Hülsken-Giesler M, Jacobs G, Kalbe E, Kerling A, Klingenberg T, Költzsch Y, Lammel-Polchau C, Ludwig W, Marschollek M, Martens B, Meis M, Meyer EM, Meyer J, Meyer Zu Schwabedissen H, Moritz N, Müller H, Nebel W, Neyer FJ, Okken PK, Rahe J,



- Remmers H, Rölker-Denker L, Schilling M, Schöpke B, Schröder J, Schulze GC, Schulze M, Siltmann S, Song B, Spehr J, Steen EE, Steinhagen-Thiessen E, Tanschus NM, Tegtbur U, Thiel A, Thoben W, van Hengel P, Wabnik S, Wegel S, Wilken O, Winkelbach S, Wist T, Wolf KH, Wolf L, Zokoll-van der Laan M; Lower Saxony Research Network GAL. Information and communication technologies for promoting and sustaining quality of life, health and self-sufficiency in ageing societies – outcomes of the Lower Saxony Research Network Design of Environments for Ageing (GAL). *Inform Health Soc Care*. 2014; 39: 166-87.
- [16] Wolf, K.-H. et al.; Studiengruppe AGT-Reha. Evaluation der Wirksamkeit und Kosten der poststationären häuslichen Tele-Rehabilitation mit AGT-Reha im Vergleich zur Medizinischen Trainingstherapie. Bericht mit Studienplan. 2016.
- [17] Steiner B, Elgert L, Haux R, Wolf KH. AGT-Reha-WK study: Protocol for a non-inferiority trial comparing the efficacy and costs of home-based telerehabilitation for shoulder diseases with medical exercise therapy. *BMJ Open*. 2020; 10: e036881.
- [18] Steiner B, Elgert L, Saalfeld B, Schwartz J, Borrmann HP, Kobelt-Pönicke A, Figlewicz A, Kasprowski D, Thiel M, Kreikebohm R, Haux R, Wolf KH. Health-enabling technologies for telerehabilitation of the shoulder: A feasibility and user acceptance study. *Methods Inf Med*. 2020; 59: e90-9.
- [19] Haux R, Karafyllis N. Methodisch-technische Aspekte der Evaluation erweiterten Zusammenwirkens. In: Haux R, Gahl K, Jipp M, Kruse R, Richter O (Hrsg.). *Zusammenwirken von natürlicher und künstlicher Intelligenz*, 175-98. Wiesbaden: Springer VS; 2021.
- [20] Gaus W, Muche R. *Medizinische Statistik*. Stuttgart: Schattauer; 2013.
- [21] van Belle G, Fisher LD, Heagerty PJ, Lumley T. *Biostatistics: A methodology for the health sciences*, 2nd edition. Hoboken, NJ: Wiley; 2004.
- [22] Rosner B. *Fundamentals of biostatistics*, 8th edition. Boston, MA: Cengage Learning; 2015.
- [23] Immich H. (1974). *Medizinische Statistik*. Stuttgart: Schattauer. Online-Fassung in der Intranet-Bibliothek der Deutschen Gesellschaft für Medizinische Informatik, Biometrie und Epidemiologie (GMDS). www.gmds.de. Zugriffen: 3. Februar 2021.
- [24] Schumacher M. Entwicklung klinischer Studien von Paul Martini bis heute. *Drug Research*. 2016; 66: 5-7.
- [25] World Medical Association (WMA). Declaration of Helsinki – ethical principles for medical research involving human subjects. Adopted 1964, last amendment 2013. <https://www.wma.net/policies-post/wma-declaration-of-helsinki-ethical-principles-for-medical-research-involving-human-subjects/>. Zugriffen: 3. Februar 2021.
- [26] Pearson K. On the criterion that a given system of derivations from the probable in the case of a correlated system of variables is such that it can be reasonably supposed to have arisen from random sampling. *The London, Edinburgh, and Dublin Philosophical Magazine and Journal of Science*. 1900; 50: 157–75.
- [27] Faul F, Erdfelder E, Lang AG, Buchner A. G*Power 3: A flexible statistical power analysis program for the social, behavioral, and biomedical sciences. *Behav Res Methods*. 2007; 39: 175-91.
- [28] Faul F, Erdfelder E, Buchner A, Lang AG. Statistical power analyses using G*Power 3.1: Tests for correlation and regression analyses. *Behav Res Methods*. 2009; 41:1149-60.
- [29] Lenhard W, Lenhard A. Berechnung von Effektstärken Dettelbach: Psychometrica; 2016. <https://www.psychometrica.de/effektstaerke.html>. Zugriffen: 3. Februar 2021
- [30] Spiegelhalter D. Should We Trust Algorithms? *Harvard Data Science Review*. 2020; 2. <https://hdsr.mitpress.mit.edu/pub/56lnenzj>. Zugriffen: 3. Februar 2021.